Virtual Beach and Beach Manager Tools

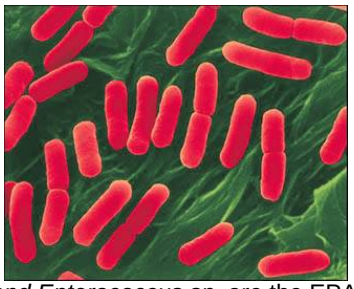
Kurt Wolfe, Walter Frick, Zhongfu Ge, Richard Zepp, Marirosa Molina, Mike Cyterski, Rajbir Parmar, Candida West

March 18, 2008

Motivation

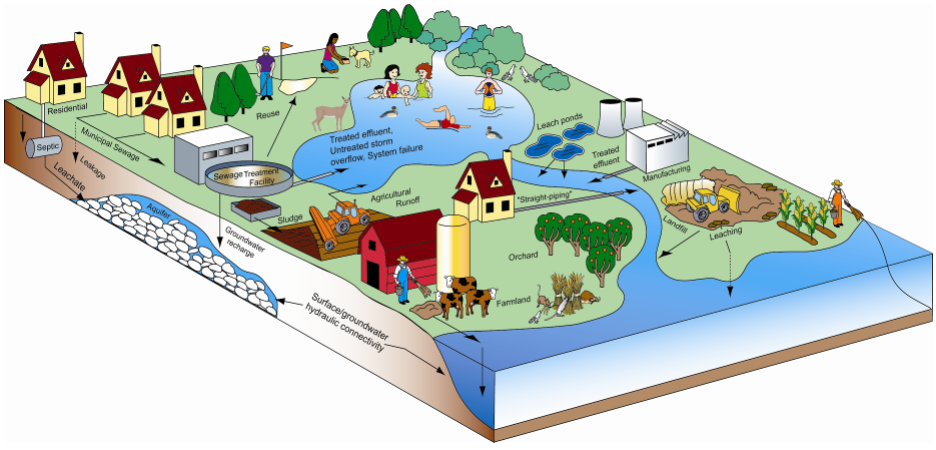
- **Clean Water Act (1972)**
  - **Recreational waters should be** "Fishable and Swimmable"
- **Beaches Environmental Assessment and Coastal Health (BEACH) Act (2000)**
  - **Provides assistance for local monitoring**
- **NRDC lawsuit**
  - **Increased pressure for action**
- **Establishment of Total Maximum Daily Loads:**
  - **Regulation that controls fecal loads into waterways**
    - ***Definition:*** **Maximum amount of a pollutant that a waterbody can receive and still meet water quality standards.**

EPA Recommended Indicators



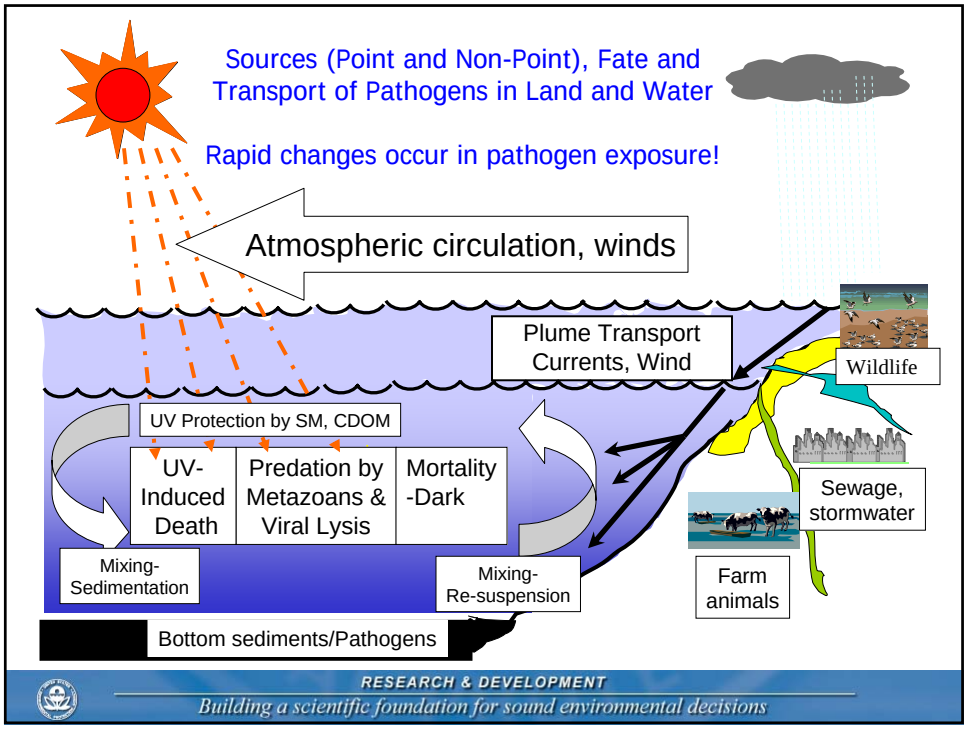
- *Escherichia coli and Enterococcus sp.* are the EPA recommended indicators to monitor for fecal contamination in water resources.
- Although pathogens are responsible for illness-reported cases, indicators are used to establish the possibility that pathogens are present.
- Sometimes pathogens and indicators do not correlate.
- Drawback: need for a minimum of 24 hr incubation time

Origins and Fate of Pathogens in the Environment



Modified from Daughton (2006), "Origin and Fate of PPCPs in the Environment"

RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions



Virtual Beach and Beach Manager Software

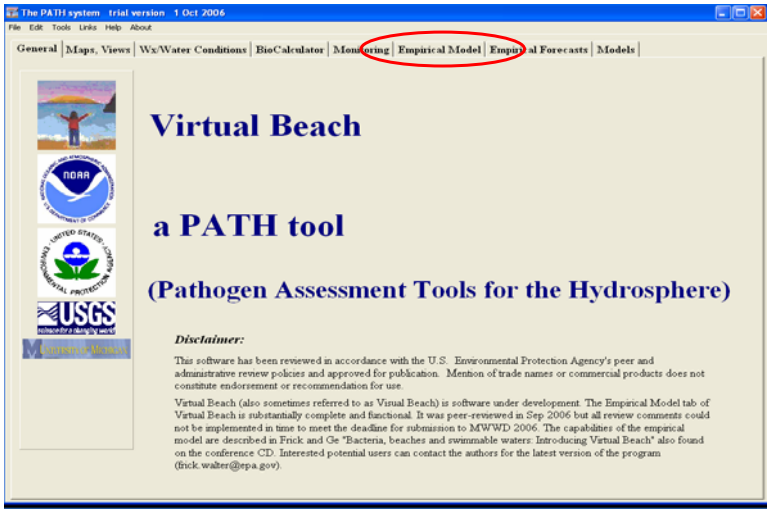
- Virtual Beach
  - Software package that facilitates developing MLR models for pathogen prediction
- Beach Manager
  - Software package provides user friendly beach advisory decision support for non technical users

RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions

Goals of the Pilot Study

- User feedback
  - Put the tools in the hands of day to day users
  - What works, what doesn't, what can we improve?
- Increase the number of sites tested
  - Varied geographic, hydrologic, meteorological conditions
- Local versus national/distributed environmental data
- Inform process related research

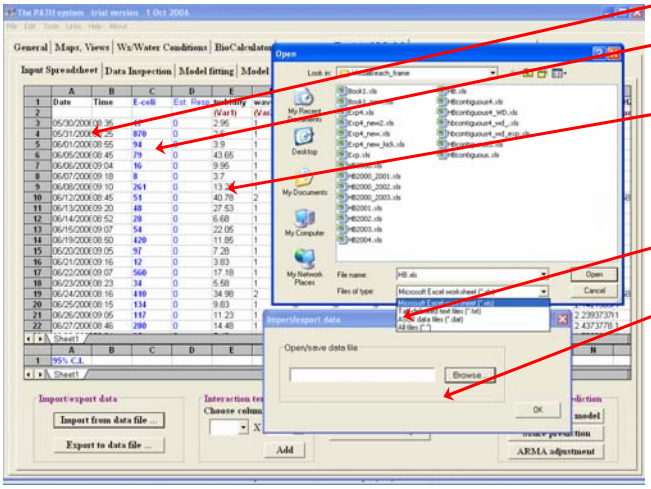
Windows application with Graphical User Interface



RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions

Virtual Beach

Compatibility with Microsoft Excel 4.0 and other major ASCII data formats

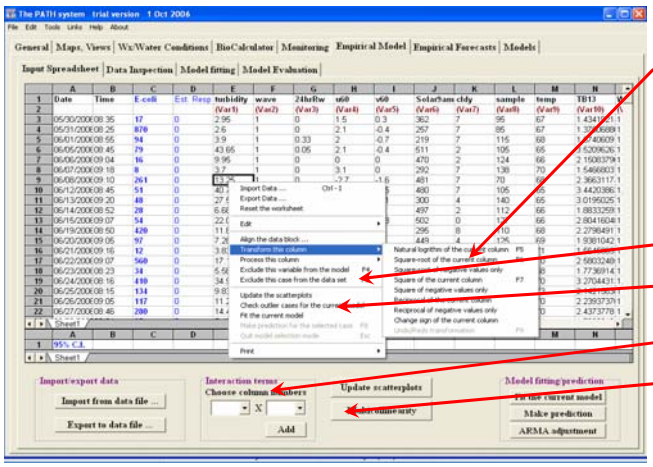


RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

Virtual Beach

Convenient data pre-processing functionalities



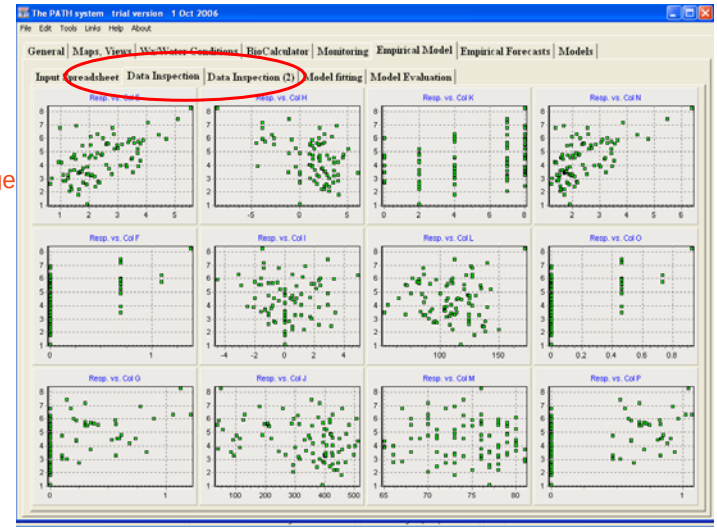
RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

Virtual Beach

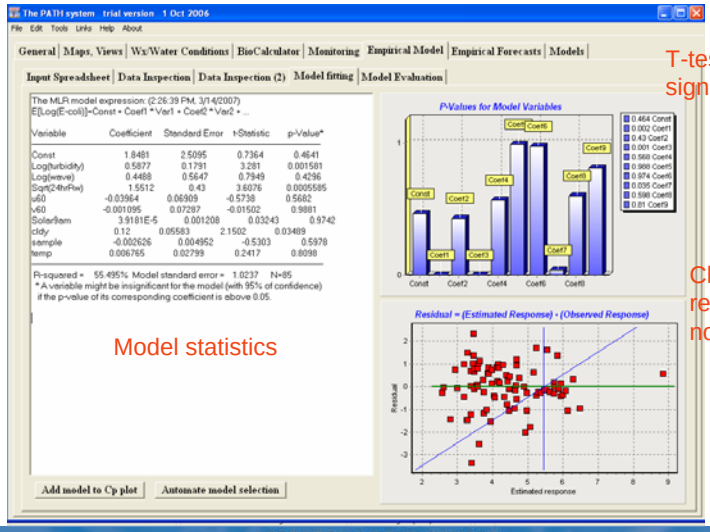
Data inspection for the robustness of the model

View multiple variable responses on one page



Virtual Beach

Finding the MLR model for the current data set



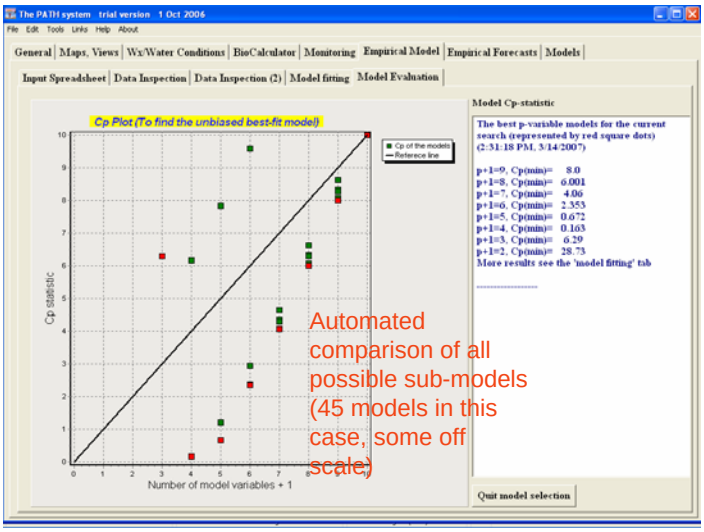
T-tests for model significance

Checking residuals for normality

Model statistics

Virtual Beach

Model assessment with Cp-plot



Automated comparison of all possible sub-models (45 models in this case, some off scale)

Virtual Beach

Graphical and text output of results

Results of the automated model selection: (2:31:18 PM, 3/14/2007)
9 variables were considered in the process:
Log(turbidity), Log(wave), Sqrt(24hrRw), u60, v60, Solar9am, cldy, sample, temp.

Model	R-squared	Stand. err.	Cp	Variable to eliminate
Full	0.555	1.024	10.0	v60
8-variable	0.555	1.017	8.0	Solar9am
7-variable	0.555	1.01	6.001	temp
6-variable	0.555	1.004	4.06	sample
5-variable	0.55	1.0	2.353	u60
4-variable	0.551	0.996	0.672	Log(wave)
3-variable	0.542	0.999	0.163	cldy
2-variable	0.494	1.044	6.29	Sqrt(24hrRw)
1-variable	0.349	1.177	28.73	N/A

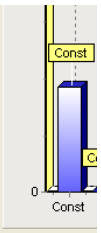
Goodness of fit

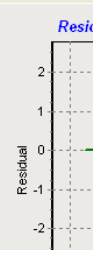
Recommended optimal model (with the min Cp):
The 3-variable model: Cp= 0.163, R-squared= 0.542
The regression equation:

E[Log(E-coli)]= 1.6923+ 0.7582 Log(turbidity)+ 1.4763 Sqrt(24hrRw)+ 0.123 cldy

(You can go back to the data sheet, eliminate variables in the squence provided above and fit the model again to obtain other regression models.)

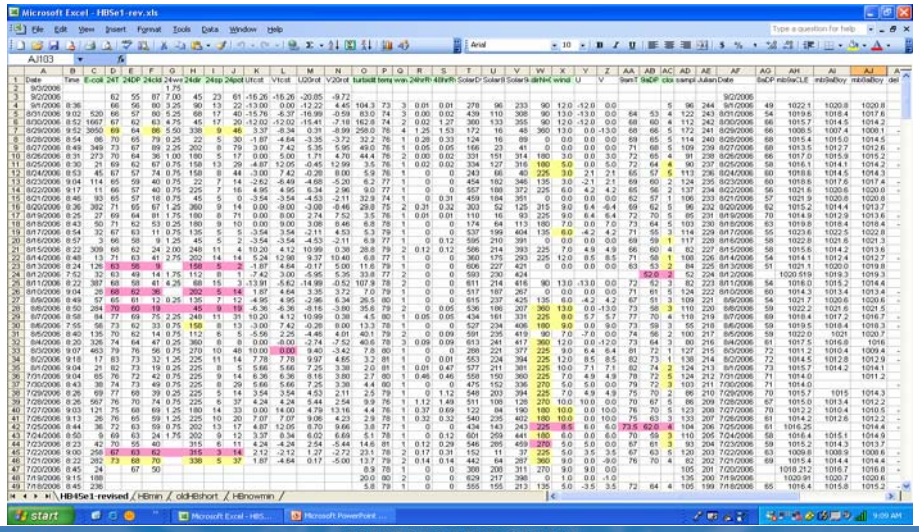
Significance of variables





Best model identified

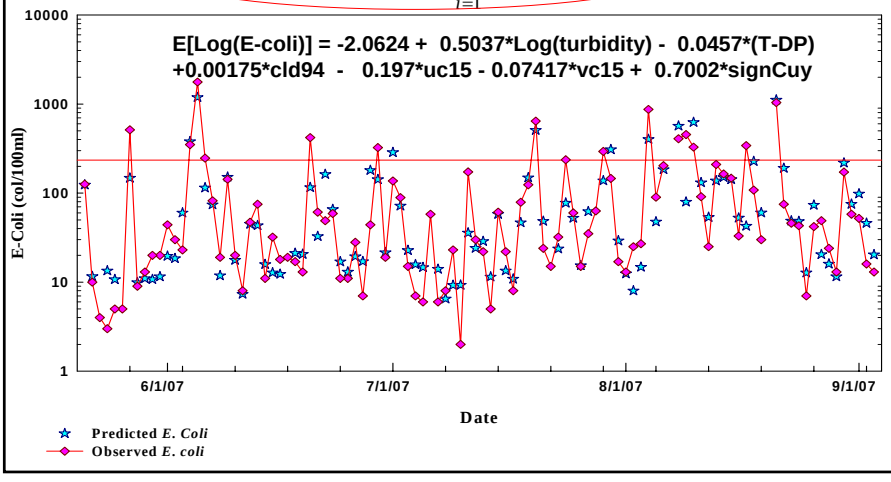
VB helps turn data like these…



… into (something like) this

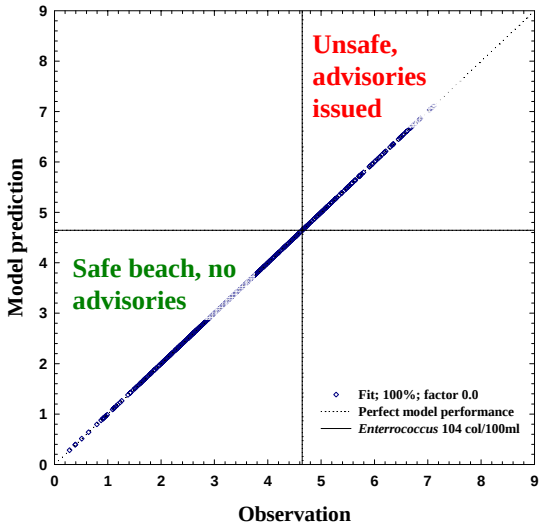
$$E[\ln(EC)] = \beta_0 + \sum_{i=1}^{p} \beta_i x_i$$

VB is not the model, but it helps select the best model!



Ideal Model

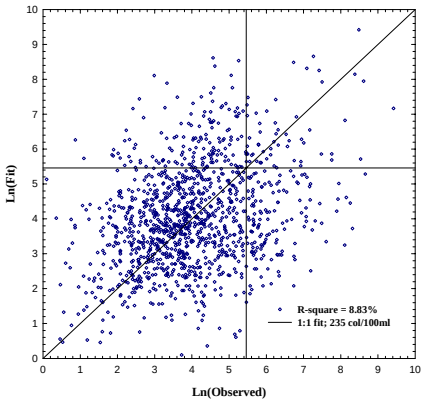
R-square = 100%



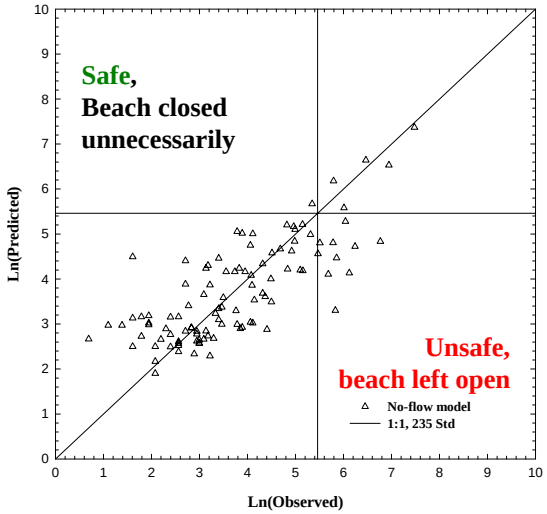
Persistence Model

With the persistence model today's concentrations are assumed to be the same as yesterday's concentrations.
R-square is 8.83%
***N* = 1184**

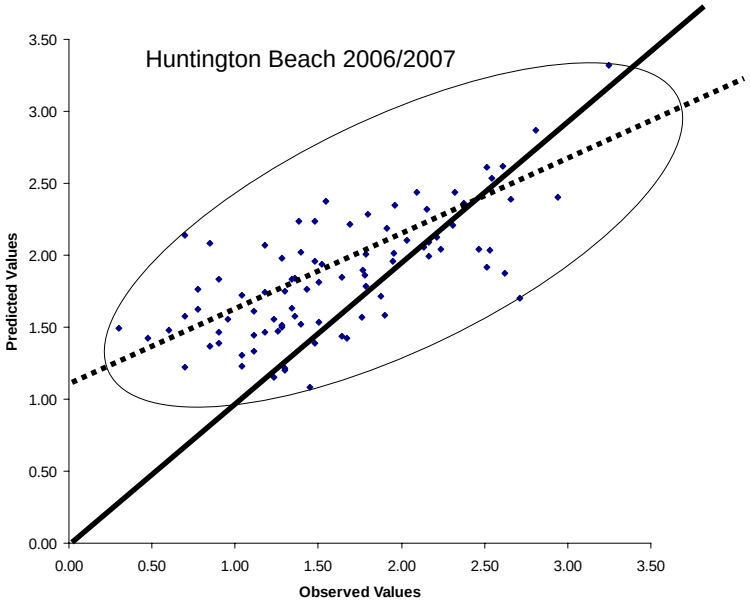
104 model
High correct: 206, 17.4%
False negatives: 159, 13.4%
False positives: 206, 13.4%
Low correct: 613, 51.8%
235 model
High correct: 57, 0.48%
False negatives: 145, 12.2%
False positives: 145, 12.2%
Low correct: 837, 70.7%



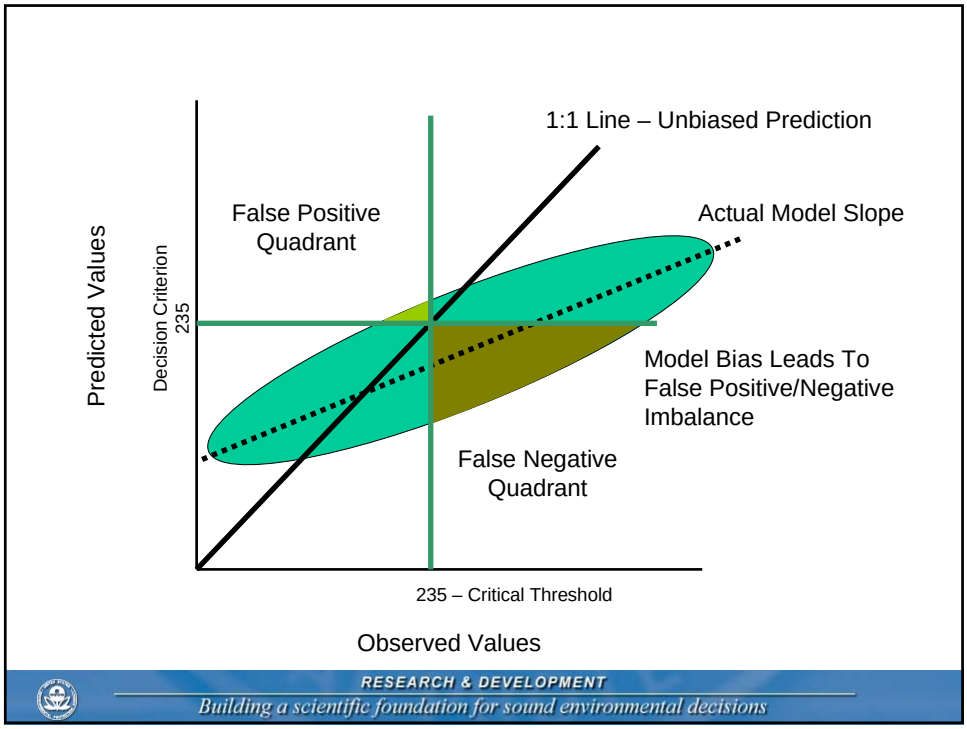
Actual Virtual Beach results

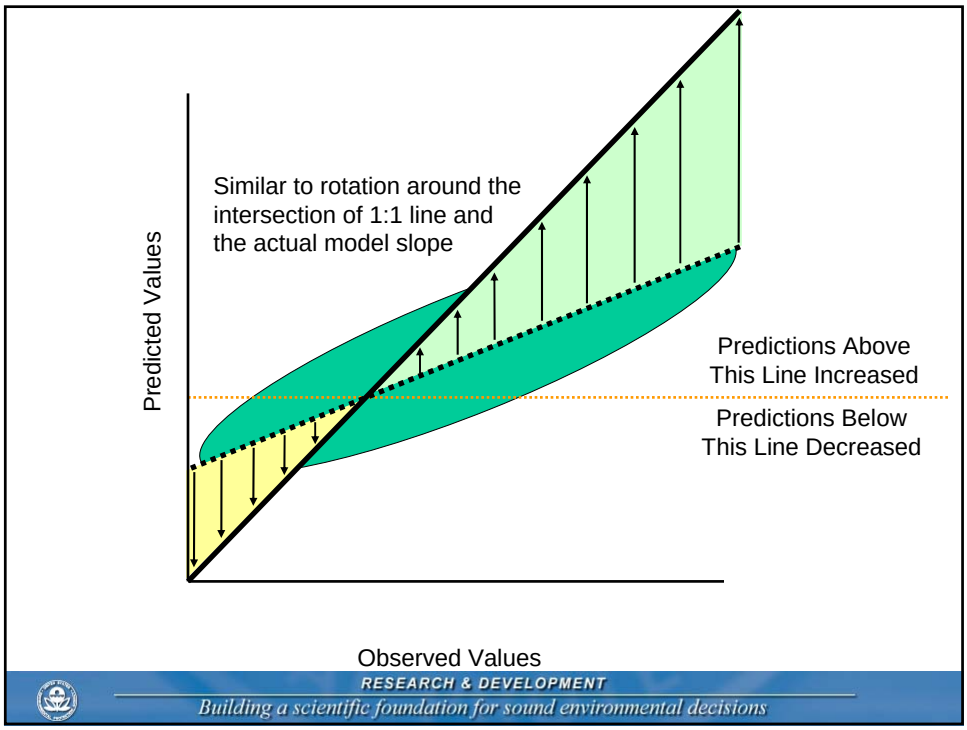


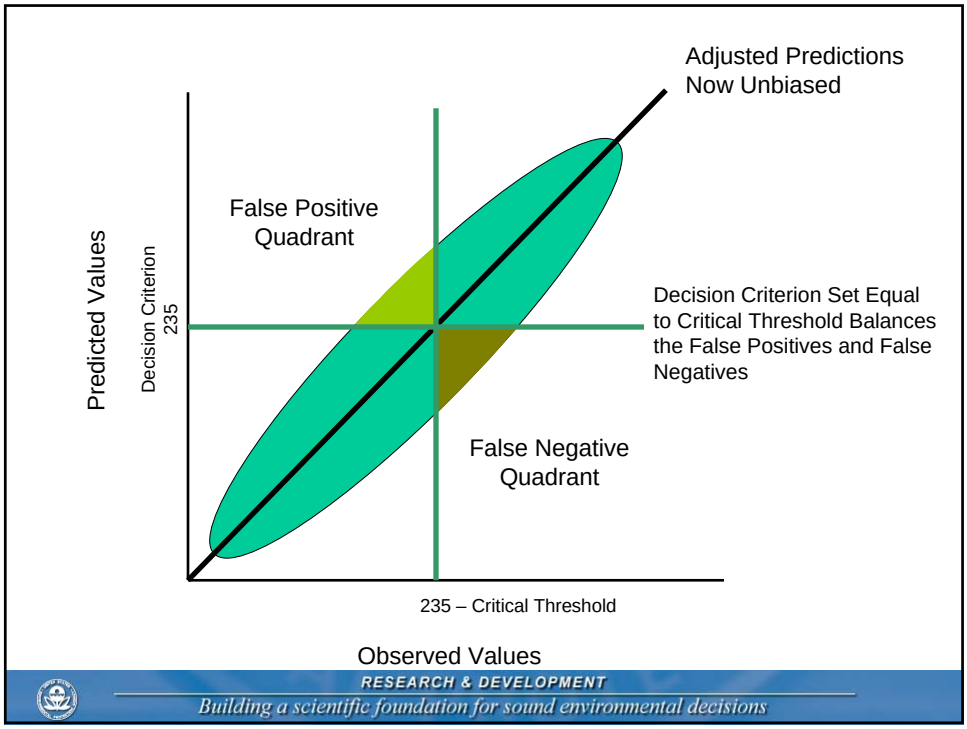
Virtual Beach model results using 2007 Huntington Beach, Ohio data.

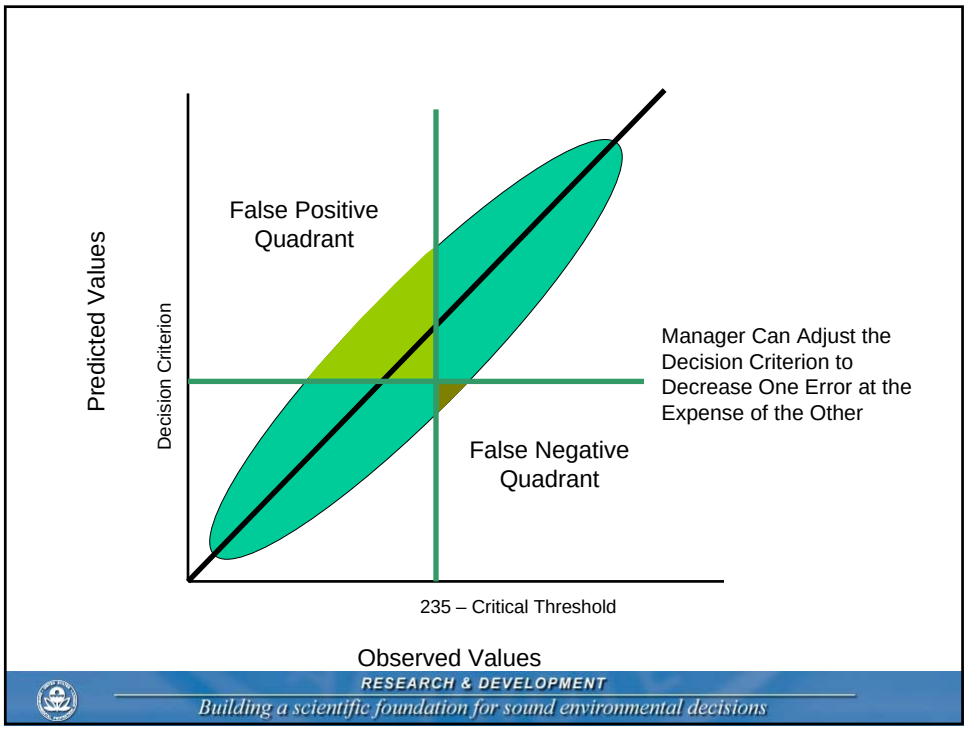


RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions


1:1 Line – Unbiased Prediction
False Positive
Quadrant
Actual Model Slope
Predicted Values
Decision Criterion
235
Model Bias Leads To
False Positive/Negative
Imbalance
False Negative
Quadrant
235 – Critical Threshold
Observed Values
RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions

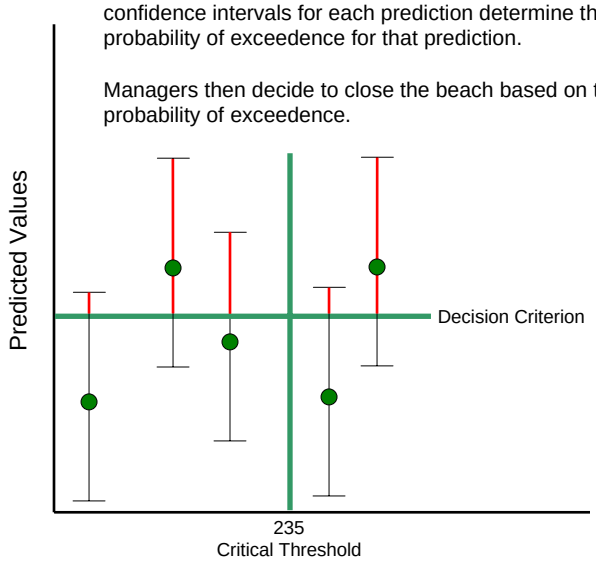

Similar to rotation around the
intersection of 1:1 line and
the actual model slope
Predicted Values
Predictions Above
This Line Increased
Predictions Below
This Line Decreased
Observed Values
RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions


Adjusted Predictions
Now Unbiased
False Positive
Quadrant
Predicted Values
Decision Criterion
235
Decision Criterion Set Equal
to Critical Threshold Balances
the False Positives and False
Negatives
False Negative
Quadrant
235 – Critical Threshold
Observed Values
RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions


False Positive
Quadrant
Predicted Values
Decision Criterion
Manager Can Adjust the
Decision Criterion to
Decrease One Error at the
Expense of the Other
False Negative
Quadrant
235 – Critical Threshold
Observed Values
RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions

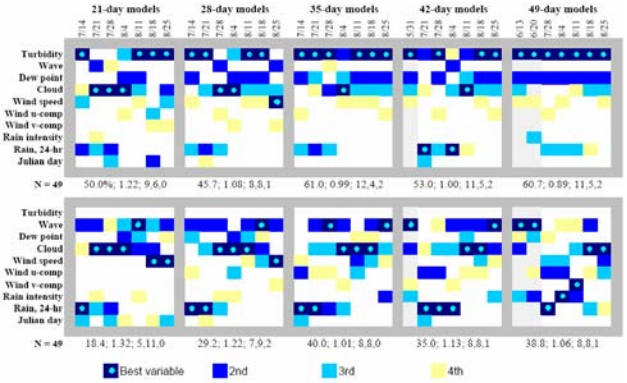
The location of the decision criterion and the (typically) 95% confidence intervals for each prediction determine the probability of exceedence for that prediction.

Managers then decide to close the beach based on this probability of exceedence.



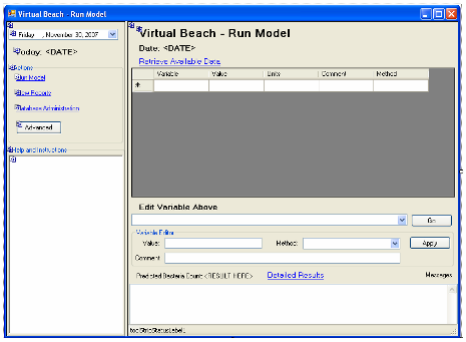
Dynamic Modeling

- Investigate use of different models over the course of a season
- Investigate use of variable data set size



Beach Manager

- Site specific MLR
  - Develop model in Virtual Beach
  - Import PMML
- Ease of use
  - Little setup required
  - MLR details hidden from user
  - Automated data download
  - Model self checking
- Data repository
  - SQLite data storage



Environmental Data Sources

- Environmental variables
  - Wind speed, wind direction, air temp, water temp, turbidity, wave height, river flow, precipitation, …
- Data sources
  - Buoys, weather stations, gauging stations, remote sensing???

Future Developments

- Investigate additional techniques
  - Neural networks
- Build more intelligence into Beach Manager
  - Automated model generation
  - Automated model selection
- Web based system
  - Centralized data access
  - Easily deployed/maintained

Contacts

Kurt Wolfe	wolfe.kurt@epa.gov
Richard Zepp	zepp.richard@epa.gov
Walter Frick	frick.walter@epa.gov
Zhongfu Ge	ge.zhongfu@epa.gov
Mike Cyterski	cyterski.mike@epa.gov
Marirosa Molina	molina.marirosa@epa.gov
Rajbir Parmar	parmar.rajbir@epa.gov
Candida West	west.candida@epa.gov

RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions